NETWORK VIDEO METHOD

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority from provisional application Serial No. 60/214,457, filed 06/30/00.

BACKGROUND OF THE INVENTION

The invention relates to electronic devices, and more particularly to video coding, transmission, and decoding/synthesis methods and circuitry.

The performance of real-time digital video systems using network transmission, such as the mobile video conferencing, has become increasingly important with current and foreseeable digital communications. Both dedicated channel and packetized-over-network transmissions benefit from compression of video signals. The widely-used motion compensation compression of video of H.263 and MPEG uses I-frames (intra frames) which are separately coded and P-frames (predicted frames) which are coded as motion vectors for macroblocks of a prior frame plus the residual difference between the motion-vector-predicted macroblocks and the actual.

Real-time video transmission over the Internet is usually done using the Real-time Transport Protocol (RTP). RTP sits on top of the User Datagram Protocol (UDP). The UDP is an unreliable protocol which does not guarantee the delivery of all the transmitted packets. Packet loss has an adverse impact on the quality of the video reconstructed at the receiver. Hence, error resilience techniques have to be adopted to mitigate the effect of packet losses. A common heuristic technique used is the frequent periodic transmission of I-frames in order to stop the propagation of errors by P-frames. That is, the motion compensation is adjusted to increase the number of I-frames and correspondingly decrease the number of P-frames.

However, this reduces the transmission rate because I-frame encoding requires many more bits than P-frame encoding.

SUMMARY OF THE INVENTION

The present invention provides a method of motion compensated video for transmission over a packetized network which trades off repeated transmission of a P-frames and the I-frame rate.

This has advantages including improved performance.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a preferred embodiment Markov chain model.

Figure 2 is a functional block diagram of a preferred embodiment encoder.

Figures 3a-3d and 4a-4d show experimental results.

Figure 5 illustrates a system.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

1. Overview

Preferred embodiment encoders and methods for motion compensated video transmission over a packetized network are illustrated generally in functional block form in Figure 2. The preferred embodiments apply a Markov chain model (illustrated in Figure 1) to control motion compensation compression by determining the rate of I-frames: a lower I-frame rate allows for repeated transmissions of P-frames as a forward error correction (FEC) method. This contrasts with the approach of increasing the I-frame rate and not repeating P-frames. In particular, the preferred embodiments maximize the probability of error-free reconstruction of frames as a function of the rate of I-frame transmission; a lower I-frame transmission rate allows for repeated transmissions of P-frames and thus increased probability of error free reception of P-frames.

2. First preferred embodiments

Figure 1 shows a Markov model for a first preferred embodiment system having two states: S_0 the state when the current video frame reconstruction has no errors and S_1 the state when the current video frame reconstruction has at least one error. The probabilities are as follows: q_0 is the probability a transmitted frame is an I-frame and $q_1 = 1-q_0$ is the probability a transmitted frame is a P-frame; B-frames are ignored for this analysis. The probability a transmitted I-frame is lost is p_{e0} and the probability a transmitted P-frame is lost is p_{e1}. Thus Figure 1 shows remaining in state S_0 with probability $q_0(1-p_{e0}) + q_1(1-p_{e1})$ which simply is the probability that an I-frame was transmitted and not lost plus the probability that a P-frame was transmitted and not lost. Similarly, the system remains in state S_1 with probability $1-q_0(1-p_{e0})$ which simply states that the only way to avoid a reconstruction error for a frame following an erroneous reconstructed frame is to receive (not lost) a transmitted I-frame because errors propagate in P-frames. Thus $q_0(1-p_{e0})$ also is the probability for transition from state S_1 to state S_0. Conversely, the probability of transition from

state S_0 to state S_1 is just the probability of losing the next frame which is simply $q_0p_{e0} + q_1p_{e1}$; that is, 1 minus the probability of remaining in state S_0. Thus the overall probability of being in state S_0 is $q_0(1-p_{e0})/(q_0 + q_1p_{e1})$ which is just the probability of an S_1 to S_0 transition divided by the sum of the probabilities of a state transition. Note that q_0 is equal to the reciprocal of the period (in frames) between I-frames; that is, if every nth frame is an I-frame, then the probability of a transmitted I-frame is 1/n.

Each transmitted packet over the Internet consists of compressed video data, an RTP header, and a UDP/IP header. Let v denote the number of bits in a packet header. For RTP/UDP/IP-based systems, v = 320. Because of this huge packet overhead, it is better to transmit as many source bits as possible in a single packet. The total size of the packet is limited by the maximum transmission unit (MTU) of the packet network. For Ethernet, the MTU is about 1500 bytes. Current Internet video applications use relatively low bitrates; and at low bitrates multiple P-frames can be fit into a single packet. A problem with transmitting multiple P-frames in a single packet is that the effect of packet loss becomes very severe because loss of a single packet leads to the loss of multiple P-frames. Hence, only one P-frame is transmitted in a packet. With an MTU of 1500 bytes, I-frames, however, do not fit into a single packet and have to be split across multiple packets. For ease of description, let:

I_0 denote the average size of an I-frame expressed in bits.

I_1 denote the average size of a P-frame in bits.

n_I denote the number of packets required for a single I-frame.

k_0 denote the total number of bits (compressed bitstream plus header bits) used to transmit an I-frame, so $k_0 = I_0 + n_Iv$ where v is the packet header size in bits.

k_1 denote the total number of bits used to transmit a P-frame.

R_T denote the maximum transmission bit rate allowed.

q_{f1} denote the number of times each P-frame is retransmitted.

Presume a constant frame rate of f frames per second. Then the bit rate of the source, R_S, can be expressed as $R_S = q_0fk_0 + q_1fk_1$ and the forward error

correction bit rate, R_F, which adds q_{f1} retransmissions of each P-frame, is $R_F = q_1 q_{f1} f k_1$ with q_{f1} nonnegative. Thus the total transmission rate, R, is $R = R_S + R_F = q_0 f k_0 + q_1 f k_1 + q_1 q_{f1} f k_1$.

Let p_e be the packet loss rate (assumed to be random) encountered on the Internet. Because only P-frames are retransmitted, the probability of loss of an I-frame is given by

$$p_{e0} = 1 - (1-p_e)^{n_I}$$

This just means that if any of the n_I packets containing a portion of an I-frame is lost, then the entire I-frame is lost. Similarly, the probability of loss of a P-frame is given by

$$p_{e1} = (1-m_1)p_e^{\lfloor qf1 \rfloor + 1} + m_1 p_e^{\lceil qf1 \rceil + 1}$$

where $\lfloor q_{f1} \rfloor$ is the largest integer not larger than q_{f1}, $\lceil q_{f1} \rceil$ is the smallest integer not smaller than q_{f1}, and m_1 is the fractional part of q_{f1}, that is, $m_1 = q_{f1} - \lfloor q_{f1} \rfloor$. Heuristically, if q_{f1} were an integer, then the probability of losing all $1+q_{f1}$ packets containing a P-frame would be the probability of losing the P-frame and so $p_{e1} = p_e^{1+qf}$. For noninteger q_{f1} the foregoing expression for p_{e1} is just the linear interpolation between integer values bracketing q_{f1}.

The preferred embodiment FEC method then determines the rate of I-frame and repeated P-frame transmissions which maximizes the probability of being in state S_0 ($=q_0(1-p_{e0})/(q_0 + q_1 p_{e1})$) given the constraint that $R \leq R_T$. Note that for a given probability of I-frame transmission, q_0, the value of q_{f1} immediately follows from taking the transmission rate $R = q_0 f k_0 + q_1 f k_1 + q_1 q_{f1} f k_1$ equal to the maximum transmission rate, R_T because f, k_0, and k_1 are fixed parameters of the system and $q_1 = 1-q_0$. Further, note that periodic transmission of I-frames implies q_0 is of the form 1/n where n is the period in frames between two I-frames and is an integer. Thus just evaluate the constrained probability of being in state S_0 for all reasonable values of n and pick the q_0 which maximizes the probability.

3. Experimental results

Two common test video sequences, "Akiyo" and "Mother and Daughter",

were used to evaluate the foregoing preferred embodiment method using the Markov model. The channel packet loss rate is assumed to be p_e = 10%. Whenever a frame or portion of a frame (in the case of an I-frame) is not received at the receiver, the evaluation simply copied the corresponding picture data from the previous frame. Note that because a large amount of data is lost with each packet loss, many of the more complicated error concealment techniques do not provide improved performance. The evaluation used two metrics: (i) average peak signal to noise ratio (PSNR) and (ii) fraction of frames reconstructed at the receiver that have a PSNR distortion of less than a threshold; the PSNR was obtained by averaging PSNR over 100 runs of transmitting the video bitstreams over a simulated packet loss channel, and the fraction of frames reconstructed for a distortion threshold t is denoted d_t.

The maximum total bitrate, R_T, was taken to be about 50 kb/s; and the quantization parameter was taken to be 8 for compressing the video sequences. For both video sequences, q_0 = 1/6 results in a bitrate around 50-55 kb/s at f = 10 frames/s; hence, the set of q_0s used was q_0 = 1/6, 1/8, ..., 1/20. Note that the source bitrate decreases as q_0 decreases. In the range q_0 = 1/6 to 1/20, q_0 = 1/6 corresponds t the case of maximum rate of transmission of I-frames. For each of the video sequences, eight bitstreams were generated, one for each value of q_0. Frame lengths l_0 and l_1 used for the Markov chain analysis were obtained by averaging the I-frame and P-frame lengths, respectively, of the compressed bitstreams; and n_I = 3 was used based on the I-frame size and MTU consideration.

For "Akiyo" the following list summarizes the parameters used for the Markov chain model:

p_e = 0.1

f = 10 frames/s

average size of I-frame, l_0 = 20,475 bits

average size of P-frame, l_1 = 1,711 bits,

R_T = 52.89 kb/s

n_I = 3

q_0 in set 1/6, 1/8, ..., 1/20

Figure 3a shows the resulting $Pr(S_0)$, the probability of being in state S_0, Figure 3b shows the average PSNR for various values of q_0, and Figure 3c shows the resulting fraction of reconstructed frames with distortion less than threshold, d_t. To obtain Figures 3b and 3c, the P-frame retransmission rate, q_{f1}, derived from the Markov chain analysis was manually tweaked so that the total bitrate (source rate + FEC rate) was very near to the source bitrate (also the total bitrate) for q_0 = 1/6. This was done to provide a fair comparison of results. Figure 3d shows the resulting total bitrate. In Figure 3d R_S denotes the source rate, R_F denotes the rate used by the FEC, and R_T denotes the total bitrate.

As can be seen from Figure 3a, the Markov chain model predicts that to obtain improved performance it makes sense to decrease the frequency of I-frames (from q_0 = 1/6 to q_0 = 1/14 .. 1/20) and to instead use retransmission of P-frames. Figures 3b and 3c support this claim. There is an improvement in average PSNR in the range of 0.4-0.55 dB and fraction of reconstructed frames which have reconstruction errors less than t, with t = 0.5, 1.0, 1.5 dB, goes up by about 0.15-0.2. The d_t curve of Figure 3c implies that there are about 20-25% more "good" frames when retransmission of P-frames is used instead of increasing the frequency of I-frame transmission.

For "Mother and Daughter" the following list summarizes the parameters used for the Markov chain model:

p_e = 0.1

f = 10 frames/s

average size of I-frame, I_0 = 18,010 bits

average size of P-frame, I_1 = 2,467 bits,

R_T = 54.84 kb/s

n_I = 3

q_0 in set 1/6, 1/8, ..., 1/20

Figure 4a shows the resulting $Pr(S_0)$, Figure 4b shows the average PSNR for various values of q_0, and Figure 4c shows the resulting d_t. To obtain Figures 4b and 4c, the P-frame retransmission rate, q_{f1}, derived from the Markov chain

analysis again was manually tweaked so that the total bitrate was very near to the source bitrate (also the total bitrate) for $q_0 = 1/6$. This was done to provide a fair comparison of results. Figure 4d shows the resulting total bitrate. In Figure 4d R_S denotes the source rate, R_F denotes the rate used by the FEC, and R_T denotes the total bitrate.

The Markov chain analysis in this case predicts that a gain in performance cannot be achieved by decreasing the frequency of I-frames; see Figure 4a. The PSNR and the d_t curves of Figure 4b and 4c support this claim. The PSNR and the d_t curves remain more or less flat. Note that the PSNR and the d_t curves do not move down like the $Pr(S_0)$ curve of Figure 4a. This can be attributed to the fact that the Markov chain model is a very simplistic model and is not based on the PSNR metric. More complex models can be thought of for modeling the PSNR performance, but they become complicated because of hte use of motion compensation in the decoder.

4. System preferred embodiments

Figure 5 shows in functional block form a portion of a preferred embodiment system which uses a preferred embodiment motion-compensated video transmission method. Such systems include video phone communication over the Internet with wireless links at the ends and voice packets interspersed with the video packets; a two-way communication version would have the structure of Figure 5 for both directions. In preferred embodiment communication systems users (transmitters and/or receivers) hardware could include one or more digital signal processors (DSP's) and/or other programmable devices such as RISC processors with stored programs for performance of the signal processing of a preferred embodiment method. Alternatively, specialized circuitry (ASIC's) could be used with (partially) hardwired preferred embodiments methods. Users may also contain analog and/or mixed-signal integrated circuits for amplification or filtering of inputs to or outputs from a communications channel and for conversion between analog and digital. Such analog and digital circuits may be integrated on a single die. The stored programs, including codebooks,

may, for example, be in ROM or flash EEPROM or FeRAM which is integrated with the processor or external to the processor. Antennas may be parts of receivers with multiple finger RAKE detectors for air interface to networks such as the Internet. Exemplary DSP cores could be in the TMS320C6xxx and TMS320C5xxx families from Texas Instruments.

5. Modifications

The preferred embodiments may be modified in various ways while retaining one or more of the features of optimization of I-frame rate in view of repeated P-frame transmission possibilities.

For example, the predictively-coded frames could include B-frames; the frame playout could include a large buffer and delay to allow from some automatic repeat request for I-frame packets to supersede some repeat P-frame packets; the network protocols could differ.

Indeed, one can introduce the concept of using multiple servers to serve the same video receiving client. For example, presume the use of two video servers to serve the same client. This situation has two network channels feeding into the video client. Use one channel to transmit the I-frame and P-frame (without repetition) and then use the other channel to transmit the FEC P-frames. Note that the rate of video received at the client is the same as when a single server is used. Use of two channels improves the performance, because the probability of both the channels deteriorating at the same time decreases.